

A Frequency Domain Test for Propriety of Complex-Valued Vector Time Series

Swati Chandna and Andrew T. Walden, *Senior Member, IEEE*

Abstract—This paper proposes a frequency domain approach to test the hypothesis that a complex-valued vector time series is proper, i.e., for testing whether the vector time series is uncorrelated with its complex conjugate. If the hypothesis is rejected, frequency bands causing the rejection will be identified and might usefully be related to known properties of the physical processes. The test needs the associated spectral matrix which can be estimated by multitaper methods using, say, K tapers. Standard asymptotic distributions for the test statistic are of no use since they would require $K \rightarrow \infty$, but, as K increases so does resolution bandwidth which causes spectral blurring. In many analyses K is necessarily kept small, and hence our efforts are directed at practical and accurate methodology for hypothesis testing for small K . Our generalized likelihood ratio statistic combined with exact cumulant matching gives very accurate rejection percentages and outperforms other methods. We also prove that the statistic on which the test is based is comprised of canonical coherencies arising from our complex-valued vector time series. Our methodology is demonstrated on ocean current data collected at different depths in the Labrador Sea.

Overall this work extends results on propriety testing for complex-valued vectors to the complex-valued vector time series setting.

Index Terms—Generalized likelihood ratio test (GLRT), multichannel signal, spectral analysis.

I. INTRODUCTION

There has long been an interest in time series motions on the complex plane: the rotary analysis method decomposes such motions into counter-rotating components which have proved particularly useful in the study of geophysical flows influenced by the rotation of the Earth [7], [8], [19], [32], [33].

Let a complex-valued p -vector-valued discrete time series be denoted $\{\mathbf{Z}_t\}$. This has as t -th element, ($t \in \mathbb{Z}$), the column vector $\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{p,t}]^T$. A length- N realization of $\{\mathbf{Z}_t\}$ namely z_0, \dots, z_{N-1} has $z_t \in \mathbb{C}^p$. In this paper we assume the p processes are jointly second-order stationary.

We propose a frequency domain approach to testing the hypothesis that a complex-valued p -vector-valued time series is proper, i.e., for testing whether the vector time series $\{\mathbf{Z}_t\}$ is uncorrelated with its complex conjugate $\{\mathbf{Z}_t^*\}$. If we denote the covariance sequence between these terms by $\{r_{\mathbf{Z},\tau}\}$ then propriety corresponds to $r_{\mathbf{Z},\tau} = \mathbf{0}$ for all $\tau \in \mathbb{Z}$, or $R_{\mathbf{Z}}(f) = \mathbf{0}$ over the Nyquist frequency range, where

$R_{\mathbf{Z}}(f)$ is the Fourier transform of $\{r_{\mathbf{Z},\tau}\}$. Otherwise the time series is said to be improper; the practical importance and occurrence of improper processes is discussed in, e.g., [1], [22] and [28].

The relevance of propriety for two-component complex-valued series ($p = 2$) can be found in [19]. Because the series are complex, two types of cross-covariance can be defined: that between the two series, known as the inner cross-covariance [19], and that between one series and the complex conjugate of the other, known as the outer cross-covariance [19]. If the vector time series is proper then the outer cross-covariance is everywhere zero.

In this paper we take as an example a six-component complex-valued ocean current time series recorded in the Labrador Sea. Frequency domain analysis is particularly useful in a scientific setting: if the hypothesis is rejected, frequency bands causing the rejection can be identified and quite possibly related to known properties of the physical processes.

Analogous tests applicable to complex-valued random vectors — rather than time series — have been described by, e.g., [29] and [34]. However, we need to consider new methodology suitable for very limited degrees of freedom. Our test uses the associated spectral matrix which can be estimated by multitaper methods using, say, K tapers. Standard asymptotic distributions for the test statistic are of no use since they would require $K \rightarrow \infty$, but, as K increases so does resolution bandwidth which causes spectral blurring. In many analyses K is necessarily kept small, and hence our efforts are directed at practical and accurate methodology for hypothesis testing for small K . Our generalized likelihood ratio statistic combined with exact cumulant matching gives very accurate rejection percentages and outperforms competitor methods.

For the scalar case, ($p = 1$), a parametric hypothesis test for propriety of complex time series is given in [30], [31]. This is based on the series being well-modelled by a Matérn process in [30] or complex autoregressive process of order one in [31], and utilises the χ^2 distribution for the test statistic, an asymptotic result. This is in contrast to our approach which (i) is suitable for $p > 1$, (ii) is nonparametric, so does not rely on a good fit to a parametric model, and (iii) develops a suitable non-asymptotic distribution for the test statistic.

Our test statistic is comprised of canonical coherencies arising from the complex-valued vector time series, analogous to the situation for complex-valued random vectors. Canonical analysis of real-valued vector time series has been extensively studied and utilised (e.g., [20], [26]), mostly in the context of parametric autoregressive moving-average (ARMA) models. Miyata [21] looked at real-valued vector time series, and

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. S. Chandna is with Dept. of Statistical Science, University College London, London WC1E 6BT, UK (e-mail: s.chandna@ucl.ac.uk). A. T. Walden is with the Dept. of Mathematics, Imperial College London, London SW7 2AZ, UK (e-mail: a.walden@imperial.ac.uk).

developed canonical correlations through linear functions of discrete vector Fourier transforms of two sets of time series. Rather than work with the Fourier transforms, which are sample values, we instead work with the orthogonal processes underlying the complex-valued vector time series, and whose variances and cross-covariances correspond exactly to the spectral components. We are thus able to define population — as well as sample — canonical coherencies for complex-valued vector time series.

Our methodology is demonstrated on ocean current data collected at different depths in the Labrador Sea.

A. Contributions

Following some background in Section II on complex-valued time series, and the statistical properties of their spectral matrix estimators under the Gaussian stationary assumption for $\{\mathbf{Z}_t\}$, the contributions of this paper are as follows:

- 1) In Section III we formally derive the canonical coherencies for $\{\mathbf{Z}_t\}$ and $\{\mathbf{Z}_t^*\}$ and show in Section IV how a test statistic $T(f)$ for testing $\mathbf{R}_Z(f) = \mathbf{0}$ arises from the sample canonical coherencies.
- 2) After giving further research context in Section V, we carefully study the statistical properties of $M(f) = -2K \log T(f)$ in Section VI, concentrating on the small K case. We show that Box's scaled chi-square approximation is exact for $p = 1$ but not for $p > 1$, and we derive the cumulants of $M(f)$.
- 3) In Section VII we show that for $p > 1$ and small K matching the first three cumulants of $M(f)$ exactly to a scaled F distribution performs at least as well as competitor methods.
- 4) A simulation study is given in Section VIII which supports the use of the scaled F approximation for $M(f)$ for the complex-valued vector time series setting. A data analysis using 6-vector valued oceanographic time series is given in Section IX which shows that when propriety is rejected, the frequency domain approach usefully shows which frequency bands cause the rejection, which may be linked to the physical processes involved.
- 5) In Section X we show how our use of canonical coherencies in the complex-valued setting is quite different to an existing approach in the literature derived for real-valued processes, even though there are some structural features in common.

II. BACKGROUND

A. Some Definitions

We consider a complex-valued p -vector-valued discrete time stochastic process $\{\mathbf{Z}_t\}$ whose t th element, $t \in \mathbb{Z}$, is the column vector $\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{p,t}]^T$, and without loss of generality take each component process to have zero mean. The sample interval is Δ_t and the Nyquist frequency is $f_N = 1/(2\Delta_t)$. We assume the p processes are jointly second-order stationary (SOS), i.e., $\text{cov}\{Z_{l,t+\tau}, Z_{m,t}\} = E\{Z_{l,t+\tau} Z_{m,t}^*\}$ and $\text{rel}\{Z_{l,t+\tau}, Z_{m,t}\} = E\{Z_{l,t+\tau} Z_{m,t}\}$, $1 \leq l, m \leq p$, are functions of τ only. Note that $\text{rel}\{Z_{l,t+\tau}, Z_{m,t}\} =$

$\text{cov}\{Z_{l,t+\tau}, Z_{m,t}^*\}$, the covariance between one process and the complex conjugate of the other.

A matrix autocovariance sequence is then given by $s_{\mathbf{Z},\tau} = E\{\mathbf{Z}_{t+\tau} \mathbf{Z}_t^H\}$, $\tau \in \mathbb{Z}$, where superscript H denotes Hermitian (complex-conjugate) transpose. We define $s_{\mathbf{Z},lm,\tau} \equiv (s_{\mathbf{Z},\tau})_{lm}$, and a matrix cross-relation sequence follows as $r_{\mathbf{Z},\tau} = E\{\mathbf{Z}_{t+\tau} \mathbf{Z}_t^T\}$, $\tau \in \mathbb{Z}$, with $r_{\mathbf{Z},lm,\tau} \equiv (r_{\mathbf{Z},\tau})_{lm}$. From their definitions we see that

$$s_{\mathbf{Z},lm,\tau} = s_{\mathbf{Z},ml,-\tau}^*; \quad r_{\mathbf{Z},lm,\tau} = r_{\mathbf{Z},ml,-\tau}, \quad 1 \leq l, m \leq p.$$

We assume $\sum_{\tau=-\infty}^{\infty} |s_{\mathbf{Z},lm,\tau}| < \infty$ and $\sum_{\tau=-\infty}^{\infty} |r_{\mathbf{Z},lm,\tau}| < \infty$, for $1 \leq l \leq m \leq p$, which means that the Fourier transforms $S_{\mathbf{Z},lm}(f)$ and $R_{\mathbf{Z},lm}(f)$ for $1 \leq l, m \leq p$, exist and are bounded and continuous. In fact for $|f| \leq f_N$, the corresponding matrices are defined as

$$\begin{aligned} S_{\mathbf{Z}}(f) &= \Delta_t \sum_{\tau=-\infty}^{\infty} s_{\mathbf{Z},\tau} e^{-i2\pi f \tau \Delta_t} \quad \text{and} \\ R_{\mathbf{Z}}(f) &= \Delta_t \sum_{\tau=-\infty}^{\infty} r_{\mathbf{Z},\tau} e^{-i2\pi f \tau \Delta_t}. \end{aligned}$$

We note that

$$r_{\mathbf{Z},\tau} = r_{\mathbf{Z},-\tau}^T \implies R_{\mathbf{Z}}(f) = R_{\mathbf{Z}}^T(-f), \quad (1)$$

a result which will prove useful later.

The covariance stationarity means that there exists [36, p. 317] an orthogonal process $\mathbf{Z}(f)$ such that

$$\mathbf{Z}_t = \int_{-1/2}^{1/2} e^{i2\pi f t} d\mathbf{Z}(f)$$

where

$$E\{\mathbf{Z}(f') \mathbf{Z}^H(f)\} = \begin{cases} S_{\mathbf{Z}}(f) df, & f = f' \\ 0, & \text{otherwise.} \end{cases}$$

B. Proper Processes

If $r_{\mathbf{Z},\tau} = \mathbf{0}$ for all $\tau \in \mathbb{Z}$, or $R_{\mathbf{Z}}(f) = \mathbf{0}$ for all $|f| \leq f_N$, then the process $\{\mathbf{Z}_t\}$ is said to be *proper*. Equivalently we see that if $\{\mathbf{Z}_t\}$ is uncorrelated with its complex conjugate $\{\mathbf{Z}_t^*\}$, then the vector-valued process is proper. This paper considers the problem of testing that the vector process is proper.

Remark 1: Based on the naming convention adopted in [28, p. 41] for complex-valued vectors, an alternative would be to call the component processes ‘jointly proper.’

C. Spectral Matrices

Let

$$Z_{l,t} = X_{l,t} + iY_{l,t}, \quad (2)$$

with $\{X_{l,t}\}$ and $\{Y_{l,t}\}$ real-valued, for $l = 1, \dots, p$, where $\mathbf{V}_t = [\mathbf{X}_t^T, \mathbf{Y}_t^T]^T = [X_{1,t}, \dots, X_{p,t}, Y_{1,t}, \dots, Y_{p,t}]^T$ is a real $2p$ -dimensional vector-valued Gaussian stationary process. Then if

$$\mathbf{T} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{I}_p & i\mathbf{I}_p \\ \mathbf{I}_p & -i\mathbf{I}_p \end{bmatrix}, \quad (3)$$

we see that

$$\mathbf{T} \mathbf{V}_t = \begin{bmatrix} \mathbf{X}_t + i\mathbf{Y}_t \\ \mathbf{X}_t - i\mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_t^* \end{bmatrix} = \mathbf{U}_t, \quad (4)$$

where $\mathbf{U}_t = [\mathbf{Z}_t^T, \mathbf{Z}_t^H]^T = [Z_{1,t}, \dots, Z_{p,t}, Z_{1,t}^*, \dots, Z_{p,t}^*]^T$ is a real $2p$ -dimensional vector-valued Gaussian stationary process.

The spectral matrix for \mathbf{V}_t is given by

$$\mathbf{S}_V(f) = \begin{bmatrix} \mathbf{S}_{XX}(f) & \mathbf{S}_{XY}(f) \\ \mathbf{S}_{YX}(f) & \mathbf{S}_{YY}(f) \end{bmatrix} \in \mathbb{C}^{2p \times 2p}. \quad (5)$$

The spectral matrix for \mathbf{U}_t is $\mathbf{S}_U(f) = \mathbf{T} \mathbf{S}_V(f) \mathbf{T}^H$ and has the form

$$\mathbf{S}_U(f) = \begin{bmatrix} \mathbf{S}_Z(f) & \mathbf{R}_Z(f) \\ \mathbf{R}_Z^H(f) & \mathbf{S}_Z^*(-f) \end{bmatrix} \in \mathbb{C}^{2p \times 2p}. \quad (6)$$

The matrix $\mathbf{S}_U(f)$ can be written in the alternative covariance matrix form

$$E\{\mathbf{U}(f) \mathbf{U}^H(f)\} = \mathbf{S}_U(f) df,$$

where

$$\mathbf{U}(f) \stackrel{\text{def}}{=} [d\mathbf{Z}^T(f), d\mathbf{Z}^H(-f)]^T. \quad (7)$$

D. Estimation

Given a length- N sample $\mathbf{V}_0, \dots, \mathbf{V}_{N-1}$, form $h_{k,t} \mathbf{V}_t$ using a suitable set of K length- N orthonormal data taper sequences $\{h_{k,t}\}, k = 0, \dots, K-1$, and compute

$$\mathbf{J}_{V,k}(f) = \Delta_t^{1/2} \sum_{t=0}^{N-1} h_{k,t} \mathbf{V}_t e^{-i2\pi f t \Delta_t}.$$

In this work we use sine tapers (e.g., [35]).

As $N \rightarrow \infty$, with the number of degrees of freedom, K fixed, and with the given taper properties, $\{\mathbf{J}_{V,k}(f), k = 0, 1, \dots, K-1\}$ are proper, independent and identically distributed random vectors such that

$$\mathbf{J}_{V,k}(f) \stackrel{d}{=} \mathcal{N}_{2p}^C(\mathbf{0}, \mathbf{S}_V(f)), \quad 0 < |f| < f_N, \quad (8)$$

for $k = 0, \dots, K-1$ (e.g., [4]). As $\mathbf{J}_{U,k}(f) = \mathbf{T} \mathbf{J}_{V,k}(f)$, as $N \rightarrow \infty$, with K fixed, $\{\mathbf{J}_{U,k}(f), k = 0, 1, \dots, K-1\}$ are also a set of proper, independent and identically distributed random vectors each of which are distributed as

$$\mathbf{J}_{U,k}(f) \stackrel{\text{def}}{=} \mathcal{N}_{2p}^C(\mathbf{0}, \mathbf{S}_U(f)), \quad 0 < |f| < f_N. \quad (9)$$

The probability density function (PDF) of $\mathbf{J}_{U,k}(f)$ — a proper Gaussian vector in \mathbb{C}^{2p} is given by [24]

$$\pi^{-p} [\det\{\mathbf{S}_U(f)\}]^{-1} \exp \left\{ -\mathbf{J}_{U,k}^H(f) \mathbf{S}_U^{-1}(f) \mathbf{J}_{U,k}(f) \right\}. \quad (10)$$

The independence of $\mathbf{J}_{U,k}(f)$'s allows us to write the joint PDF of $\mathbf{J}_{U,0}(f), \dots, \mathbf{J}_{U,K-1}(f)$ as the product of their marginal densities given by (10). So the likelihood function, $g_J(\mathbf{S}_U(f) | \mathbf{J}_{U,0}(f), \dots, \mathbf{J}_{U,K-1}(f))$, of $\mathbf{S}_U(f)$ given $\mathbf{J}_{U,0}(f), \dots, \mathbf{J}_{U,K-1}(f)$, is given by

$$[\pi^p \det\{\mathbf{S}_U(f)\}]^{-K} \exp \left\{ -\sum_{k=0}^{K-1} \mathbf{J}_{U,k}^H(f) \mathbf{S}_U^{-1}(f) \mathbf{J}_{U,k}(f) \right\}. \quad (11)$$

Now $\hat{\mathbf{S}}_U(f)$ is the sample covariance matrix of $\{\mathbf{J}_{U,k}(f); k = 0, 1, \dots, K-1\}$, i.e.,

$$\hat{\mathbf{S}}_U(f) = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{J}_{U,k}(f) \mathbf{J}_{U,k}^H(f) = \begin{bmatrix} \hat{\mathbf{S}}_Z(f) & \hat{\mathbf{R}}_Z(f) \\ \hat{\mathbf{R}}_Z^H(f) & \hat{\mathbf{S}}_Z^*(-f) \end{bmatrix}. \quad (12)$$

Noting that the argument of $\exp\{\cdot\}$ in (11) is scalar, and so is equal to its trace, and recalling the linearity and cyclicity of the trace operator, we can write

$$g_J = [\pi^p \det\{\mathbf{S}_U(f)\}]^{-K} \exp \left\{ -K \text{tr}\{\mathbf{S}_U^{-1}(f) \hat{\mathbf{S}}_U(f)\} \right\}, \quad (13)$$

where dependence of g on its arguments is suppressed for convenience.

For a finite value of N , $\{\mathbf{J}_{U,k}(f); k = 0, 1, \dots, K-1\}$ are proper random variables with

$$\mathbf{J}_{U,k}(f) \stackrel{d}{=} \mathcal{N}_{2p}^C(\mathbf{0}, \mathbf{S}_U(f)), \quad W_N < |f| < f_N - W_N, \quad (14)$$

where $[-W_N, W_N]$ is the extent of the spectral window induced by tapering [4]. For sine tapers

$$W_N = (K+1)/[2(N+1)\Delta_t], \quad (15)$$

(e.g., [35]). Therefore, in practice, we have to restrict interest to frequencies in the range $W_N < |f| < f_N - W_N$.

III. CANONICAL COHERENCIES

The structure of the testing problem is related to measures of coherence between vector-valued processes, and so we next turn our attention to the idea of canonical coherence.

A. New Series Defined by Cross-correlations

Consider the cross-correlation of complex-valued deterministic matrix sequence $\{\mathbf{A}_t\}$ with the time series $\{\mathbf{Z}_t\}$ to give $\{\boldsymbol{\xi}_t\}$:

$$\boldsymbol{\xi}_t = \mathbf{A}^* \star \mathbf{Z}_t \stackrel{\text{def}}{=} \sum_{u=-\infty}^{\infty} \mathbf{A}_u^* \mathbf{Z}_{t+u}.$$

Likewise we define the cross-correlation of complex-valued deterministic matrix sequence $\{\mathbf{B}_t\}$ with the time series $\{\mathbf{Z}_t^*\}$ to give $\{\boldsymbol{\eta}_t\}$:

$$\boldsymbol{\eta}_t = \mathbf{B}^* \star \mathbf{Z}_t^* \stackrel{\text{def}}{=} \sum_{u=-\infty}^{\infty} \mathbf{B}_u^* \mathbf{Z}_{t+u}^*.$$

Component-wise we have

$$\begin{bmatrix} \xi_{1,t} \\ \xi_{2,t} \\ \vdots \\ \xi_{p,t} \end{bmatrix} = \sum_u \begin{bmatrix} a_{11,u}^* & \cdots & \cdots & a_{1p,u}^* \\ a_{21,u}^* & \cdots & \cdots & a_{2p,u}^* \\ \vdots & & & \vdots \\ a_{p1,u}^* & \cdots & \cdots & a_{pp,u}^* \end{bmatrix} \begin{bmatrix} Z_{1,t+u} \\ Z_{2,t+u} \\ \vdots \\ Z_{p,t+u} \end{bmatrix}. \quad (16)$$

So, for $j = 1, \dots, p$,

$$\xi_{j,t} = \sum_u a_{j1,u}^* Z_{1,t+u} + \cdots + \sum_u a_{jp,u}^* Z_{p,t+u}. \quad (17)$$

The spectral representation theorem allows us to write $\xi_{j,t}$, $j = 1, \dots, p$ and $Z_{l,t}$, $l = 1, \dots, p$, as

$$\xi_{j,t} = \int_{-f_N}^{f_N} e^{i2\pi f t \Delta_t} dZ_{\xi_j}(f); \quad Z_{l,t} = \int_{-f_N}^{f_N} e^{i2\pi f t \Delta_t} dZ_l(f).$$

Substituting the spectral representation for $Z_{1,t}$ in the first term of (17), we get

$$\sum_u a_{j1,u}^* Z_{1,t+u} = \int_{-f_N}^{f_N} e^{i2\pi f t \Delta_t} A_{j1}^*(f) dZ_1(f),$$

where $A_{jl}(f) = \sum_u a_{jl,u} e^{-i2\pi f u \Delta_t}$. Proceeding in analogous fashion, and using the fact that the orthogonal process in a spectral representation is unique [6, p. 34], we obtain

$$\begin{aligned} dZ_{\xi_j}(f) &= A_{j1}^*(f) dZ_1(f) + \dots + A_{jp}^*(f) dZ_p(f) \\ &\stackrel{\text{def}}{=} \mathbf{A}_j^H(f) d\mathbf{Z}(f). \end{aligned}$$

So

$$\xi_{j,t} = \int_{-f_N}^{f_N} e^{i2\pi f t \Delta_t} \mathbf{A}_j^H(f) d\mathbf{Z}(f). \quad (18)$$

For $\{\eta_t\}$ a similar procedure gives

$$\begin{aligned} dZ_{\eta_j}(f) &= B_{j1}^*(f) dZ_1^*(-f) + \dots + B_{jp}^*(f) dZ_p^*(-f) \\ &\stackrel{\text{def}}{=} \mathbf{B}_j^H(f) d\mathbf{Z}^*(-f), \end{aligned}$$

and

$$\eta_{j,t} = \int_{-f_N}^{f_N} e^{i2\pi f t \Delta_t} \mathbf{B}_j^H(f) d\mathbf{Z}^*(-f). \quad (19)$$

The usual definition of the (magnitude squared) coherencies $\gamma_j^2(f)$ between series $\{\xi_{j,t}\}$ and $\{\eta_{j,t}\}$ is

$$\begin{aligned} \gamma_j^2(f) &= \frac{|E\{dZ_{\xi_j}(f) dZ_{\eta_j}^H(f)\}|^2}{E\{|dZ_{\xi_j}(f)|^2\} E\{|dZ_{\eta_j}(f)|^2\}} \\ &= |\text{corr}\{dZ_{\xi_j}(f), dZ_{\eta_j}(f)\}|^2. \end{aligned}$$

Remark 2: It should be emphasized that throughout we use the usual definition of coherence as a magnitude squared quantity, basically a squared correlation coefficient.

B. Finding Canonical Coherencies

In vector notation,

$$d\mathbf{Z}_\xi(f) = \mathbf{A}^H(f) d\mathbf{Z}(f) \text{ and } d\mathbf{Z}_\eta(f) = \mathbf{B}^H(f) d\mathbf{Z}^*(-f), \quad (20)$$

where $\mathbf{A}(f) = [\mathbf{A}_1(f), \mathbf{A}_2(f), \dots, \mathbf{A}_p(f)]$.

Consider $|\text{corr}\{dZ_{\xi_j}(f), dZ_{\eta_j}(f)\}|$. This can be written

$$\begin{aligned} &\frac{|\mathbf{A}_j^H(f) \mathbf{R}_Z(f) d\mathbf{f} \mathbf{B}_j(f)|}{[\mathbf{A}_j^H(f) \mathbf{S}_Z(f) d\mathbf{f} \mathbf{A}_j(f)]^{1/2} [\mathbf{B}_j^H(f) \mathbf{S}_Z^T(-f) d\mathbf{f} \mathbf{B}_j(f)]^{1/2}} \\ &= \frac{|\mathbf{A}_j^H(f) \mathbf{R}_Z(f) \mathbf{B}_j(f)|}{[\mathbf{A}_j^H(f) \mathbf{S}_Z(f) \mathbf{A}_j(f)]^{1/2} [\mathbf{B}_j^H(f) \mathbf{S}_Z^T(-f) \mathbf{B}_j(f)]^{1/2}}. \end{aligned} \quad (22)$$

Suppose we choose $\mathbf{A}(f)$ and $\mathbf{B}(f)$ so that

$$\mathbf{A}^H(f) \mathbf{S}_Z(f) \mathbf{A}(f) = \mathbf{I}_p = \mathbf{B}^H(f) \mathbf{S}_Z^T(-f) \mathbf{B}(f). \quad (21)$$

Then

$$|\text{corr}\{dZ_{\xi_j}(f), dZ_{\eta_j}(f)\}| = |\mathbf{A}_j^H(f) \mathbf{R}_Z(f) \mathbf{B}_j(f)|.$$

It also ensures that for $j \neq k$,

$$\text{corr}\{dZ_{\xi_j}(f), dZ_{\xi_k}(f)\} = 0 = \text{corr}\{dZ_{\eta_j}(f), dZ_{\eta_k}(f)\}. \quad (22)$$

Define

$$\mathbf{K}(f) \stackrel{\text{def}}{=} \mathbf{A}^H(f) \mathbf{R}_Z(f) \mathbf{B}(f),$$

so that

$$|K_{jj}(f)| = |\text{corr}\{dZ_{\xi_j}(f), dZ_{\eta_j}(f)\}|.$$

Definition 1: The first definition of the canonical coherence problem under the standardization in (21) is as follows. Find $\mathbf{A}_1(f)$ and $\mathbf{B}_1(f)$ such that $|K_{11}(f)| = |\text{corr}\{dZ_{\xi_1}(f), dZ_{\eta_1}(f)\}|$ is maximized. Next find $\mathbf{A}_2(f)$ and $\mathbf{B}_2(f)$ such that $|K_{22}(f)| = |\text{corr}\{dZ_{\xi_2}(f), dZ_{\eta_2}(f)\}|$ is maximized, subject to $dZ_{\xi_2}(f), dZ_{\eta_2}(f)$ being uncorrelated with $dZ_{\xi_1}(f), dZ_{\eta_1}(f)$. In general, at step j for $j = 2, \dots, p$, $\mathbf{A}_j(f)$ and $\mathbf{B}_j(f)$ are found such that $|K_{jj}(f)| = |\text{corr}\{dZ_{\xi_j}(f), dZ_{\eta_j}(f)\}|$ is maximized subject to $dZ_{\xi_j}(f), dZ_{\eta_j}(f)$ being uncorrelated with $dZ_{\xi_k}(f), dZ_{\eta_k}(f)$ for $1 \leq k < j$.

The problem can be defined in a different but equivalent way [27].

Definition 2: The second definition of the canonical coherence problem under the standardization in (21) is as follows. Choose $\mathbf{A}(f)$ and $\mathbf{B}(f)$ such that all partial sums over the $|K_{jj}(f)|$ are maximized, i.e.,

$$\max_{\mathbf{A}(f), \mathbf{B}(f)} \sum_{j=1}^r |K_{jj}(f)|, \quad r = 1, \dots, p. \quad (23)$$

Lemma 1: The canonical coherencies

$$l_j^2(f) \stackrel{\text{def}}{=} |K_{jj}(f)|^2, \quad j = 1, \dots, p$$

and $\mathbf{A}_j(f)$ and $\mathbf{B}_j(f)$ for $j = 1, \dots, p$, solving (23) are eigenvalues and eigenvectors defined as follows:

$$\begin{aligned} \mathbf{S}_Z^{-1}(f) \mathbf{R}_Z(f) \mathbf{S}_Z^{-T}(-f) \mathbf{R}_Z^H(f) \mathbf{A}_j(f) &= l_j^2(f) \mathbf{A}_j(f) \\ \mathbf{S}_Z^{-T}(-f) \mathbf{R}_Z^H(f) \mathbf{S}_Z^{-1}(f) \mathbf{R}_Z(f) \mathbf{B}_j(f) &= l_j^2(f) \mathbf{B}_j(f). \end{aligned}$$

Moreover we have that as a result,

$$\text{corr}\{dZ_{\xi_j}(f), dZ_{\eta_k}(f)\} = 0, \quad \text{for } j, k = 1, \dots, p; j \neq k. \quad (24)$$

Proof: See Appendix A. ■

Remark 3: From Lemma 1 the optimal $\mathbf{A}_j(f)$ and $\mathbf{B}_j(f)$ give rise to the j th pair of canonical series via (18) and (19).

Remark 4: Results (22) and (24) ensure that the uncorrelated requirements in Definition 1 hold.

IV. GENERALIZED LIKELIHOOD RATIO TEST

A. Formulation

The GLR test statistic for

$$H_0 : \mathbf{R}_Z(f) = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{R}_Z(f) \neq \mathbf{0}, \quad (25)$$

for any $W_N < |f| < f_N - W_N$, is given by ratio of the likelihood function (13) with $\mathbf{S}_U(f)$ constrained to have zero off-diagonal blocks ($\mathbf{R}_Z(f) = \mathbf{0}$) to the likelihood function with $\mathbf{S}_U(f)$ unconstrained, i.e.,

$$\frac{\max_{\mathbf{S}_U(f): \mathbf{R}_Z(f)=\mathbf{0}} g_J}{\max_{\mathbf{S}_U(f)} g_J} \stackrel{\text{def}}{=} L_G(f). \quad (26)$$

The unconstrained maximum likelihood estimate of the covariance matrix $\mathbf{S}_U(f)$ is given by the corresponding sample covariance matrix $\hat{\mathbf{S}}_U(f)$ in (12), thus maximum likelihood estimate of $\mathbf{S}_U(f)$ under the constraint $\mathbf{R}_Z(f) = \mathbf{0}$ is,

$$\check{\mathbf{S}}_U(f) = \begin{bmatrix} \hat{\mathbf{S}}_Z(f) & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{S}}_Z^T(-f) \end{bmatrix}. \quad (27)$$

From (13), (26) it follows that $T(f) \stackrel{\text{def}}{=} L_G^{1/K}(f)$ is

$$\begin{aligned} T(f) &= \frac{[\det\{\check{\mathbf{S}}_U(f)\}]^{-1} \exp\{-\text{tr}\{\check{\mathbf{S}}_U^{-1}(f)\hat{\mathbf{S}}_U(f)\}\}}{[\det\{\hat{\mathbf{S}}_U(f)\}]^{-1} \exp\{-\text{tr}\{\hat{\mathbf{S}}_U^{-1}(f)\hat{\mathbf{S}}_U(f)\}\}} \\ &= \det\{\check{\mathbf{S}}_U^{-1}(f)\hat{\mathbf{S}}_U(f)\} \\ &\times \exp\left\{-\text{tr}\{\check{\mathbf{S}}_U^{-1}(f)\hat{\mathbf{S}}_U(f) - \mathbf{I}_{2p}\}\right\}. \end{aligned} \quad (28)$$

The result (13) is valid for $W_N < |f| < f_N - W_N$, but from (1) we see that if $\mathbf{R}_Z(f) = \mathbf{0}$ for $f > 0$ then it is also $\mathbf{0}$ for $f < 0$. Hence in practice we need only concern ourselves with the positive frequency range $W_N < f < f_N - W_N$, and calculate $T(f)$ over this interval.

From (12) and (27) we see that

$$\check{\mathbf{S}}_U^{-1}(f)\hat{\mathbf{S}}_U(f) = \begin{bmatrix} \mathbf{I}_p & \hat{\mathbf{S}}_Z^{-1}(f)\hat{\mathbf{R}}_Z(f) \\ \hat{\mathbf{S}}_Z^{-T}(-f)\hat{\mathbf{R}}_Z^H(f) & \mathbf{I}_p \end{bmatrix},$$

so that the $\exp\{\cdot\}$ term is unity. Thus (28) becomes

$$T(f) = \det\{\check{\mathbf{S}}_U^{-1}(f)\hat{\mathbf{S}}_U(f)\} \quad (29)$$

$$\begin{aligned} &= \det\left\{\begin{bmatrix} \mathbf{I}_p & \hat{\mathbf{S}}_Z^{-1}(f)\hat{\mathbf{R}}_Z(f) \\ \hat{\mathbf{S}}_Z^{-T}(-f)\hat{\mathbf{R}}_Z^H(f) & \mathbf{I}_p \end{bmatrix}\right\} \\ &= \det\{\mathbf{I}_p - \hat{\mathbf{S}}_Z^{-1}(f)\hat{\mathbf{R}}_Z(f)\hat{\mathbf{S}}_Z^{-T}(-f)\hat{\mathbf{R}}_Z^H(f)\} \quad (30) \\ &= \frac{\det\{\hat{\mathbf{S}}_Z(f) - \hat{\mathbf{R}}_Z(f)\hat{\mathbf{S}}_Z^{-T}(-f)\hat{\mathbf{R}}_Z^H(f)\}}{\det\{\hat{\mathbf{S}}_Z(f)\}}. \end{aligned}$$

Starting with (29) and using (27) we also have that

$$T(f) = \frac{\det\{\hat{\mathbf{S}}_U(f)\}}{\det\{\check{\mathbf{S}}_U(f)\}} = \frac{\det\{\hat{\mathbf{S}}_U(f)\}}{\det\{\hat{\mathbf{S}}_Z(f)\}\det\{\hat{\mathbf{S}}_Z^T(-f)\}}. \quad (31)$$

Now, the GLR test may be based on any of the above equivalent forms for $T(f)$. Form (31), unlike other formulations does not involve computation of either $\hat{\mathbf{S}}_Z^{-1}(f)$ or $\hat{\mathbf{S}}_Z^{-T}(-f)$.

By definition of the GLR test statistic (26), we shall reject the null hypothesis of $\mathbf{R}_Z(f) = \mathbf{0}$, for small values of $T(f)$. For a given size α , the rule is to reject \mathbf{H}_0 iff

$$T(f; N, K, p) \leq c, \quad (32)$$

where $\Pr(T(f; N, K, p) \leq c | H_0) = \alpha$. Here we have used the more precise notation $T(f; N, K, p)$ which emphasizes the dependence of the GLR test on (i) the sample size N , (ii) the number of tapers K (also the number of complex degrees of freedom), and (iii) dimension p of the complex time series.

B. Invariance

Now

$$\mathbf{R}_Z(f)df \stackrel{\text{def}}{=} E\{d\mathbf{Z}(f)d\mathbf{Z}^T(-f)\}.$$

Apply $\mathbf{L}(f) \in \mathbb{C}^{p \times p}$ to $d\mathbf{Z}(f)$ so that $d\mathbf{Z}(f) \rightarrow \mathbf{L}(f)d\mathbf{Z}(f)$, and therefore $d\mathbf{Z}^T(-f) \rightarrow \mathbf{L}^*(-f)d\mathbf{Z}^T(-f)$. So

$$\begin{aligned} \mathbf{R}_Z(f) &= \mathbf{0} \implies E\{\mathbf{L}(f)d\mathbf{Z}(f)[\mathbf{L}^*(-f)d\mathbf{Z}^T(-f)]^H\} \\ &= \mathbf{L}(f)\mathbf{R}_Z(f)d\mathbf{f}\mathbf{L}^T(-f) = \mathbf{0}, \end{aligned}$$

i.e., $\mathbf{R}_Z(f) = \mathbf{0}$ is invariant to the linear transformation $d\mathbf{Z}(f) \rightarrow \mathbf{L}(f)d\mathbf{Z}(f)$. So the decision rule for our GLR test must be likewise invariant.

Note that under this transformation,

$$\begin{aligned} \mathbf{U}(f) &\rightarrow [\mathbf{L}(f)d\mathbf{Z}(f), \mathbf{L}^*(-f)d\mathbf{Z}^*(-f)]^T \\ &= \begin{bmatrix} \mathbf{L}(f) & \mathbf{0} \\ \mathbf{0} & \mathbf{L}^*(-f) \end{bmatrix} \mathbf{U}(f) \\ &\stackrel{\text{def}}{=} \mathbf{Q}(f)\mathbf{U}(f), \end{aligned}$$

so that we require invariance under the group action $\mathbf{S}_U(f) \rightarrow \mathbf{Q}(f)\mathbf{S}_U(f)\mathbf{Q}^H(f)$.

Under the null hypothesis $\mathbf{S}_U(f)$ takes the form in (27) so that $\mathbf{Q}(f)\mathbf{S}_U(f)\mathbf{Q}^H(f)$ is

$$\begin{bmatrix} \mathbf{L}(f)\mathbf{S}_Z(f)\mathbf{L}^H(f) & \mathbf{0} \\ \mathbf{0} & \mathbf{L}^*(-f)\mathbf{S}_Z^*(-f)\mathbf{L}^T(-f) \end{bmatrix},$$

and the choice $\mathbf{L}(f) = \mathbf{S}_Z^{-1/2}(f)$ (which exists for $\mathbf{S}_Z(f)$ positive definite) renders the matrix equal to \mathbf{I}_{2p} . This means that under the null hypothesis we can always replace $\mathbf{S}_U(f)$ by \mathbf{I}_{2p} without loss of generality.

From Lemma 1 we know that the eigenvalues $\ell_j^2(f)$ of $\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)$ are canonical coherencies which are invariant under the group action specified above; moreover, the corresponding empirical or sample canonical coherencies are maximal invariant and the GLR statistic — which requires this invariance — must be a function of them.

Let $\ell_j^2(f)$, $j = 1, \dots, p$, be the sample versions of the canonical coherencies $\ell_j^2(f)$ between $d\mathbf{Z}(f)$ and $d\mathbf{Z}^*(-f)$. They are the sample eigenvalues of $\hat{\mathbf{S}}_Z^{-1}(f)\hat{\mathbf{R}}_Z(f)\hat{\mathbf{S}}_Z^{-T}(-f)\hat{\mathbf{R}}_Z^H(f)$. Then from (30) it follows that for $W_N < f < f_N - W_N$,

$$T(f) = \det(\mathbf{I}_p - \hat{\mathbf{S}}_Z^{-1}(f)\hat{\mathbf{R}}_Z(f)\hat{\mathbf{S}}_Z^{-T}(-f)\hat{\mathbf{R}}_Z^H(f)) \quad (33)$$

$$= \prod_{j=1}^p (1 - \ell_j^2(f)). \quad (34)$$

V. RESEARCH CONTEXT

Testing $\mathbf{R}_Z(f) = \mathbf{0}$ is the same as testing the independence of two complex Gaussian p -vectors, namely $d\mathbf{Z}(f)$ and $d\mathbf{Z}^*(-f)$, (see (7)). The GLR test based on (31) falls in the class of multiple independence tests in multivariate statistics theory. Some distributional results for the complex case were given in [14] but did not include the case of interest here, namely two p -vectors. A later paper [9] gave the exact distribution of a power of $T(f)$ but this involves an infinite sum with very complicated components; small K

approximations were not discussed. Other relevant results can be found in [12] and [15], and these are discussed in detail in Section VII-A.

The statistic $T(f)$ is the frequency-domain time series analogue to those used in [23], [29] and [34] to examine independence between a Gaussian random vector and its complex conjugate. In [23], [29] a complex formulation was maintained but only an asymptotic approach to testing was considered. In [34] a real-valued representation of the problem was used and Box's scaled chi-square method was used to improve on the asymptotic critical values. In the rest of this paper we adopt the complex formulation, derive Box's refinement, but also improve on it for $p > 1$ by *exactly* matching the first three cumulants to a scaled F -distribution. (We point out that Box's refinement is exact for $p = 1$.) This latter F -method is very simple to implement practically, involving only the first three polygamma functions.

We emphasize that our efforts are directed at practical and accurate methodology for small K . This is important in a time series setting where as K increases so does resolution bandwidth which potentially causes spectral blurring. In many analyses K must necessarily be kept small. In the remainder of this paper we will always assume any frequency under consideration to lie in the interval $W_N < f < f_N - W_N$.

VI. BASIC PROPERTIES OF TEST STATISTIC

A. Asymptotic Behaviour

The application of Wilk's theorem [37, p. 132] gives that under H_0 , as $K \rightarrow \infty$,

$$M(f) \stackrel{\text{def}}{=} -2 \log L_G(f) = -2K \log T(f) \xrightarrow{d} \chi_\nu^2 \quad (35)$$

where \xrightarrow{d} denotes convergence in distribution and χ_ν^2 denotes the chi-square distribution with ν degrees of freedom. Here ν is the difference between the number of free real parameters under H_0 and H_1 . Comparing $\check{\mathbf{S}}_U(f)$ in (27) (for H_0) and $\mathbf{S}_U(f)$ in (6) (for H_1) we note that $\mathbf{R}_Z^H(f)$ follows directly from $\mathbf{R}_Z(f)$ so that there is only an additional $2p^2$ degrees of freedom, i.e., those contributed by $\mathbf{R}_Z(f)$. Hence we have $\nu = 2p^2$.

While (35) is a very useful and convenient result when the exact distribution of the GLR test statistic is analytically intractable, K here denotes the number of tapers used for multitaper spectral estimation and not the sample size N . For a given value of N , K could be around 10 or less. Since (35) is an asymptotic result, K must be sufficiently large to expect a reasonable χ_ν^2 approximation to $-2K \log T(f)$. Since K may not be large in a time series setting, a small- K approximation to the distribution of the test statistic under the null hypothesis is imperative.

B. Moments

Since $\mathbf{J}_{U,k}(f)$, $k = 0, \dots, K-1$, are Gaussian distributed random vectors, from (12) it follows that

$$\mathbf{A} \stackrel{\text{def}}{=} K \hat{\mathbf{S}}_U(f) \stackrel{d}{=} \mathcal{W}_{2p}^C(K, \mathbf{S}_U(f)), \quad (36)$$

i.e., $\mathbf{A}(f)$ is distributed as a $2p$ -dimensional complex Wishart distribution with K complex degrees of freedom and mean $K\mathbf{S}_U(f)$. Given the form of $\mathbf{S}_U(f)$, we partition $\mathbf{A}(f)$ analogously in terms of sub-matrices as

$$\mathbf{A}(f) = \begin{bmatrix} \mathbf{A}_{11}(f) & \mathbf{A}_{12}(f) \\ \mathbf{A}_{21}(f) & \mathbf{A}_{22}(f) \end{bmatrix}. \quad (37)$$

Then the GLR test statistic in (31) can be expressed as

$$L_G^{1/K}(f) = \frac{\det\{\mathbf{A}(f)\}}{\det\{\mathbf{A}_{11}(f)\} \det\{\mathbf{A}_{22}(f)\}}. \quad (38)$$

Lemma 2: The r th moment of $L_G(f)$, namely $E\{L_G^r(f)\}$, is given by

$$\frac{\prod_{j=1}^p \Gamma(K-j+1)}{\prod_{j=1}^p \Gamma(K-j-p+1)} \frac{\prod_{j=1}^p \Gamma(K[1+r]-j-p+1)}{\prod_{j=1}^p \Gamma(K[1+r]-j+1)}. \quad (39)$$

Proof: This is given in Appendix B. \blacksquare

A random variable $0 \leq W \leq 1$ is said to be of Box-type [2, eqn. (70)] if for all $r \in \mathbb{N}$,

$$E\{W^r\} = C_0 \left[\frac{\prod_{j=1}^l b_j^{b_j}}{\prod_{i=1}^m a_i^{a_i}} \right]^r \frac{\prod_{i=1}^m \Gamma(a_i[1+r] + \vartheta_i)}{\prod_{j=1}^l \Gamma(b_j[1+r] + \zeta_j)}, \quad (40)$$

where $\sum_{i=1}^m a_i = \sum_{j=1}^l b_j$, and the constant term C_0 is

$$C_0 = \frac{\prod_{j=1}^l \Gamma(b_j + \zeta_j)}{\prod_{i=1}^m \Gamma(a_i + \vartheta_i)},$$

so that it's zero'th moment is unity.

We see that $L_G(f)$ is a random variable of Box-type with

$$m = l = p; a_i = K; b_j = K; \vartheta_i = 1 - i - p, \zeta_j = 1 - j,$$

and C_0 is

$$C_0 = \prod_{j=1}^p \frac{\Gamma(K-j+1)}{\Gamma(K-j-p+1)}.$$

C. Cumulants

The moment generating function for $M(f) = -2 \log L_G(f)$ is given by (with f suppressed), $\phi_M(s) = E\{e^{sM}\} = E\{L_G^{-2s}\}$ so using (39),

$$\phi_M(s) = C_0 \prod_{j=1}^p \frac{\Gamma(K[1-2s]-j-p+1)}{\Gamma(K[1-2s]-j+1)}.$$

The Gamma functions will be valid if $-2Ks + K - j - p + 1 > 0$ for all $j = 1, \dots, p$, which requires $-2s > (2p-1-K)/K$.

The cumulants κ_i of M can be easily obtained from the cumulant generating function by successively differentiating $\log \phi_M(s)$ and setting $s = 0$. Notice that the requirement $-2s > (2p-1-K)/K$ corresponds to $K \geq 2p$ when $s = 0$. Then, for $i \geq 1$,

$$\kappa_i = \left. \frac{d^i \log \phi_M(s)}{(ds)^i} \right|_{s=0}$$

so that κ_i is

$$[-2K]^i \sum_{j=1}^p \left[\psi^{(i-1)}(K-j-p+1) - \psi^{(i-1)}(K-j+1) \right]. \quad (41)$$

Here for $i = 1$, $\psi(x) = [d \log \Gamma(x)]/dx$ is the digamma function, while for $i = 2$ and 3 , $\psi^{(1)}(x)$ and $\psi^{(2)}(x)$ are the trigamma and tetragamma functions respectively; these are all ‘polygamma functions.’ κ_1 is the mean, κ_2 is the variance, $\kappa_3/\kappa_2^{3/2}$ is the skewness and κ_4/κ_2^2 is the excess kurtosis.

D. Scaled chi-square approximation

Box [2] provides a scaled chi-squared approximation for M of the form $M(f) \stackrel{d}{=} c_B \chi_d^2$. The constant c_B is chosen so that the cumulants of $c_B \chi_d^2$ match those of $M(f)$ up to an error of order $O(K^{-2})$. The degrees of freedom d associated with the chi-square approximation for $M(f)$ is given by Box [2]

$$\begin{aligned} d &= -2 \left[\sum_{i=1}^p \vartheta_i - \sum_{j=1}^p \zeta_j \right] \\ &= -2 \left[\sum_{i=1}^p (1 - i - p) - \sum_{j=1}^p (1 - j) \right] \\ &= -2 \left[-\sum_{i=1}^p i - \sum_{i=1}^p p + \sum_{j=1}^p j \right] = 2p^2 = \nu, \end{aligned}$$

as expected. The scaling factor c_B is a constant determined as follows [2, p. 338]. Define

$$\omega_n = \frac{(-1)^{n+1}}{n(n+1)} \left[\sum_{i=1}^p \frac{B_{n+1}(\vartheta_i)}{a_i^n} - \sum_{j=1}^p \frac{B_{n+1}(\zeta_j)}{b_j^n} \right] \quad (42)$$

where $B_n(x)$ is the Bernoulli polynomial of degree n and order unity, with

$$B_2(x) = x^2 - x + \frac{1}{6}; \quad B_3(x) = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x.$$

Subsequently, let $W_1 = 2\omega_1/d$ and $W_2 = 4\omega_2/d$, then c_B is chosen according to the following rule:

$$c_B = \begin{cases} (1 - W_1)^{-1} & \text{if } W_2 \geq W_1^2 \\ 1 + W_1 & \text{otherwise.} \end{cases}$$

Using (42) we find that

$$W_1 = \frac{p}{K}; \quad W_2 = \frac{(7p^2 - 1)}{6K^2}.$$

It is straightforward to see that $W_2 \geq W_1^2$ for all (K, p) combinations, implying that $c_B = K/(K - p)$, giving Box’s finite sample approximation as

$$M(f) \stackrel{d}{=} \frac{K}{K - p} \chi_{2p^2}^2. \quad (43)$$

(This agrees with (35) asymptotically as $K \rightarrow \infty$ for a fixed dimension p .)

For the case $p = 1$ we note that $T(f)$ in (33) becomes

$$T(f) = 1 - \frac{|\hat{R}_Z(f)|^2}{\hat{S}_Z(f)\hat{S}_Z(-f)} = 1 - \hat{\gamma}_*^2(f)$$

where $\hat{\gamma}_*^2(f)$ is the ‘conjugate coherence,’ i.e., the ordinary coherence between $\{Z_t\}$ and $\{Z_t^*\}$ (e.g., [4]). Then $M(f) =$

$-2K \log(1 - \hat{\gamma}_*^2(f))$. Under the null hypothesis it is known that

$$\hat{\gamma}_*^2(f) \stackrel{d}{=} \text{beta}(1, K - 1), \quad (44)$$

i.e., coherence has the $\text{beta}(1, K - 1)$ distribution. It then follows readily that $M(f)$ has PDF

$$f_M(x) = \frac{K - 1}{2K} e^{-x[\frac{K-1}{2K}]},$$

so that $M(f) \stackrel{d}{=} \frac{K}{K-1} \chi_2^2$ and Box’s approximation (43) is in fact *exact* for the case $p = 1$. When $p = 1$ we note that $W_2 = W_1^2$.

Remark 5: For small values of K , matching cumulants of $M(f)$ up to an error of order $O(K^{-2})$ could be problematic for $p > 1$ [2, p. 329]. This leads us to consider other approaches.

VII. OTHER STATISTICAL APPROACHES

A. Product of Independent Beta Random Variables

Lemma 3: Under the null hypothesis the distribution of $T(f)$ can be expressed as a product of independent beta random variables:

$$T(f) \stackrel{d}{=} \prod_{j=1}^p B_j, \quad (45)$$

where $B_j \stackrel{d}{=} \text{beta}(K + 1 - j - p, p)$, independently.

Proof: This is given in Appendix C. ■

Remark 6: If $p = 1$, (45) gives $T(f) \stackrel{d}{=} \text{beta}(K - 1, 1)$, as it should since $T(f) = 1 - \hat{\gamma}_*^2(f)$, and (44) holds.

In a different context Gupta [12] developed the distribution of the product of p independent beta distributions: a likelihood ratio criterion for testing a hypothesis about regression coefficients in a multivariate normal setting takes the form $\Lambda = \det\{\mathbf{V}_1\} / \det\{\mathbf{V}_1 + \mathbf{V}_2\}$ under the corresponding null hypothesis, with \mathbf{V}_1 and \mathbf{V}_2 independently distributed as

$$\mathbf{V}_1 \stackrel{d}{=} \mathcal{W}_p^C(f_1, \Sigma), \quad \mathbf{V}_2 \stackrel{d}{=} \mathcal{W}_p^C(f_2, \Sigma),$$

for integer parameters f_1, f_2 and covariance matrix Σ . Then Λ has the three-parameter complex U distribution $U(p, f_2, f_1)$ which is distributed as a product of p beta variables with $B_j \stackrel{d}{=} \text{beta}(f_1 - j + 1, f_2)$. So setting Gupta’s parameters f_1 and f_2 to $K - p$ and p , respectively, shows that $T(f)$ has the three-parameter complex U distribution $U(p, p, K - p)$. This helps only a little because there are no simple expressions for this distribution’s PDF or quantiles etc. However, by using convolution techniques Gupta did obtain some exact results for the case $p = 2$. In fact it turns out that for $p = 2$ the right-side of (43) can be improved to

$$\frac{K}{K - 2} G(1 - \alpha) \chi_8^2(1 - \alpha) \quad (46)$$

where $G(1 - \alpha)$ is an exact (tabulated) correction factor and $\chi_8^2(1 - \alpha)$ is the 100(1 - α)% point of the chi-square distribution with 8 degrees of freedom. For example for $p = 2, K = 6$ and $\alpha = (0.05, 0.01)$ the factors are (1.043, 1.051) [12, Table 1]. The work of Gupta was extended as part of [15, p. 5]

who produced tables of approximate correction factors for the right-side of (43) for $p \geq 3$ so that $M(f)$ is compared to

$$\frac{K}{K-p} G(1-\alpha) \chi_{2p}^2(1-\alpha). \quad (47)$$

Setting their parameters n and q to $K-p$ and p respectively, shows that for example for $p = 3$, $K = 8$ and $\alpha = (0.05, 0.01)$ the factors are (1.076, 1.087) [15, Table 7]. The effect of these correction factors will be discussed shortly.

Remark 7: The result (45) is very nice, and quantiles of $T(f)$ could be found through, say, successive convolution techniques, but this is very complicated — see [3], [13] who develop this approach for a related statistic.

B. Matching the first three cumulants exactly

The look-up tables of [12] and [15] are not convenient and so we now develop a simple and fast method for approximating the percentage points of the distribution of $M(f)$. Box [2] considered using the very flexible Pearson system for approximating the distribution of likelihood ratios. Box [2, p. 330] introduced a discriminant $d = (\kappa_1 \kappa_3) / (2\kappa_2^2)$, such that if $d > 1$ a Pearson type VI should be fitted; this corresponds to $W_2 > W_1^2$. For $p = 2 : 20$, $K = 1 : 100$, with $K \geq 2p$ we always found $d > 1$ using (41). (Note $p = 1$ is excluded since $W_2 = W_1^2$ in that case.)

Box [2] considered distributions of the form bF_{ν_1, ν_2} , i.e., a scaled F distribution (Pearson type VI) with parameters ν_1, ν_2 , and suggested matching cumulants *approximately*.

We have chosen to match the first three cumulants of the form (41) *exactly*; the parameters of bF_{ν_1, ν_2} are related to the cumulants via [10]

$$\begin{aligned} b &= \frac{2\kappa_1 (\kappa_1^2 \kappa_2 - \kappa_2^2 + \kappa_1 \kappa_3)}{2\kappa_1^2 \kappa_2 - 4\kappa_2^2 + 3\kappa_1 \kappa_3}, \\ \nu_1 &= \frac{4\kappa_1 (\kappa_1^2 \kappa_2 - \kappa_2^2 + \kappa_1 \kappa_3)}{4\kappa_1 \kappa_2^2 - \kappa_1^2 \kappa_3 + \kappa_2 \kappa_3}, \\ \nu_2 &= \frac{4\kappa_1^2 \kappa_2 - 8\kappa_2^2 + 6\kappa_1 \kappa_3}{\kappa_1 \kappa_3 - 2\kappa_2^2}. \end{aligned} \quad (48)$$

Then to carry out the test $M(f)$ would be compared to

$$bF_{\nu_1, \nu_2}(1-\alpha), \quad (49)$$

where $F_{\nu_1, \nu_2}(1-\alpha)$ is the $100(1-\alpha)\%$ point of the F distribution with parameters b, ν_1, ν_2 given by (48).

C. Comparison of Approximations

For some combinations of (p, K) the asymptotic result (35) is compared to Box's basic approximation (43), the adjusted Box method (46), (47) and the scaled F method (49) in Table I which gives the 95% and 99% points of the distribution of $M(f)$ according to the four approaches. There is very good agreement between the adjusted Box method and the scaled F method, the latter being quick and simple to compute. Box's basic approximation is a massive improvement on the asymptotic result. For $p = 2$ the adjusted Box approximation due to [12] is exact and we see that the scaled F approximation is therefore very accurate. Other combinations of p and small

(p, K)	Method	$\alpha = 0.05$	$\alpha = 0.01$
(2, 6)	Asymptotic	15.51	20.09
	Box	23.26	30.14
	Adjusted Box	24.26	31.67
	scaled F	24.26	31.68
(3, 8)	Asymptotic	28.87	34.81
	Box	46.19	55.69
	Adjusted Box	49.70	60.53
	scaled F	49.71	60.54
(4, 10)	Asymptotic	46.19	53.49
	Box	76.99	89.14
	Adjusted Box	84.84	99.31
	scaled F	84.85	99.30
(5, 12)	Asymptotic	67.50	76.15
	Box	115.72	130.55
	Adjusted Box	129.96	148.17
	scaled F	129.94	148.18

TABLE I
COMPARISON OF PERCENTAGE POINTS OF $M(f)$ ACCORDING TO THE ASYMPTOTIC RESULT (35), BOX'S APPROXIMATION (43), ADJUSTED BOX METHOD (46), (47) AND THE SCALED F METHOD (49).

K lead to similar results. The agreement of the scaled F approximation with the previous historically tabulated results (adjusted Box approximation) leads us to the following recommendation.

D. Recommended testing approach

In view of the discussions and results above, the following is recommended for a given choice of α :

- If $p = 1$, reject H_0 if

$$M(f) > \frac{K}{K-1} \chi_2^2(1-\alpha). \quad (50)$$

This test is distributionally exact.

- If $p \geq 2$, reject H_0 if

$$M(f) > bF_{\nu_1, \nu_2}(1-\alpha). \quad (51)$$

The accuracy of the scaled F approximation for our time series test (25) is now confirmed by simulation.

VIII. SIMULATION RESULTS

For $p \geq 2$ we will show that using the scaled F approximation test where we reject H_0 if (51) holds brings about a worthwhile accuracy improvement over Box's approximation test where we reject H_0 if

$$M(f) > \frac{K}{K-p} \chi_{2p}^2(1-\alpha). \quad (52)$$

To be able to do this we need to simulate from a model such that $\mathbf{S}_U(f)$ in (6) has $\mathbf{R}_Z(f) = \mathbf{0}$ for some frequency range. We can proceed as follows.

We know [25] that any complex second-order stationary scalar process (assumed zero mean here), whether proper or improper, can be written as the output of a widely linear filter driven by proper white noise, i.e.,

$$Z_t = \sum_{l=-\infty}^{\infty} g_l \epsilon_{t-l} + \sum_{l=-\infty}^{\infty} h_l \epsilon_{t-l}^*, \quad (53)$$

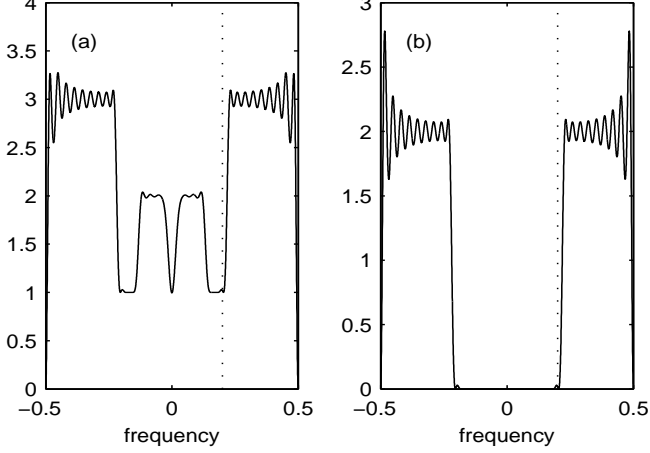


Fig. 1. (a) $S_Z(f)$ and (b) $R_Z(f)$. The vertical dotted line marks the frequency $f = 0.2$.

where $\{g_l\}$ and $\{h_l\}$ are sequence of complex constants, and $\{\epsilon_t\}$ is proper white noise for which $\text{cov}\{\epsilon_{t+\tau}, \epsilon_t\} = \sigma_\epsilon^2 \delta_{\tau,0}$ and $\text{cov}\{\epsilon_{t+\tau}, \epsilon_t^*\} = 0$, for $\tau \in \mathbb{Z}$, where $\delta_{j,k}$ is the Kronecker delta. For simulation purposes it is convenient to set $\sigma_\epsilon^2 = 1$. Then [25]

$$S_Z(f) = |G(f)|^2 + |H(f)|^2 \quad (54)$$

$$R_Z(f) = G(f)H(-f) + G(-f)H(f), \quad (55)$$

where $G(f)$ is the frequency response function of $\{g_l\}$ given by $G(f) = \sum_{l=-\infty}^{\infty} g_l e^{-i2\pi f l}$ and $H(f)$ is the frequency response function of $\{h_l\}$.

For $p \geq 2$ we generate processes $\{Z_{j,t}\}$, $j = 1, \dots, p$, such that

$$Z_{j,t} = \sum_{l=-\infty}^{\infty} g_l \epsilon_{j,t-l} + \sum_{l=-\infty}^{\infty} h_l \epsilon_{j,t-l}^* \quad (56)$$

$$+ \sum_{l=-\infty}^{\infty} a_l \bar{\epsilon}_{j,t-l} + \sum_{l=-\infty}^{\infty} a_l \bar{\epsilon}_{j,t-l}^*, \quad (57)$$

where the $2p$ processes $\{\{\epsilon_{j,t}\}, \{\bar{\epsilon}_{j,t}\}, j = 1, \dots, p\}$ are all independent of each other. The filter $\{g_l\}$ was chosen to be low-pass with a frequency transition zone $[0.125, 0.15]$. The filter $\{h_l\}$ was of ‘Hilbert-type’ or all-pass in the frequency zone $[0.05, 0.45]$. Thus $G(f)$ is real and symmetric while $H(f)$ is imaginary and skew-symmetric. According to (55), if using just these two filters, the resulting $R_Z(f)$ is zero for $f \in [-0.5, 0.5]$. However, the filter $\{a_l\}$ was chosen to be high-pass above $f = 0.2$ and therefore generates non-zero $R_Z(f)$ values at these high frequencies. The resulting $S_Z(f)$ and $R_Z(f)$ are shown in Fig. 1.

The matrix $\mathbf{S}_Z(f)$ is thus of the form $\mathbf{S}_Z(f) = S_Z(f) \mathbf{I}_p$ with frequency dependence as shown in Fig. 1(a) while $\mathbf{R}_Z(f)$ is of the form $\mathbf{R}_Z(f) = R_Z(f) \mathbf{I}_p$ with frequency dependence as shown in Fig. 1(b). We can thus simulate from this model to evaluate our hypothesis tests, knowing that for frequencies where $\mathbf{R}_Z(f) = \mathbf{0}$ in fact $\mathbf{S}_Z(f) \neq \mathbf{0}$ and thus (31) is well-defined.

Sample results are shown in Table II for $(p, K) = (2, 6)$ and $(3, 8)$. So here $K = 6$ and 8 are indeed small. Here $N = 512$

(p, K)	$100\alpha\%$	f				
		0.06	0.12	0.18	0.24	0.42
(2, 6)	1%	1.5	1.5	1.4	31.1	35.4
		1.1	1.1	0.9	25.8	30.0
	5%	6.1	6.2	6.3	60.0	63.6
(3, 8)		5.0	5.1	5.2	55.3	59.3
	1%	2.0	2.1	2.2	55.7	58.9
		0.9	1.1	1.1	42.3	45.5
	5%	8.2	8.3	8.3	81.0	82.9
		4.9	5.1	5.2	72.6	75.2

TABLE II
REJECTION PERCENTAGES OVER 10 000 REPETITIONS. THE TOP LINE OF EACH ENTRY IS FOR BOX’S χ^2 APPROXIMATION (43) AND THE LOWER LINE IS FOR THE F APPROXIMATION OF (51).

but smaller time series lengths such as 128 produced very similar results. Shown are rejection percentages for H_0 over 10 000 independent repetitions. The nominal rates are shown in the second column. The first three columns of rejection percentages are for frequencies where $\mathbf{R}_Z(f) = \mathbf{0}$, (H_0 is true) and the latter two are for frequencies where $\mathbf{R}_Z(f) \neq \mathbf{0}$ (H_0 is false) — see Fig. 1(b). The top line of each entry is for Box’s χ^2 approximation (52) and the lower line is for the F approximation of (51). We see that, proportionately, the latter has a much more accurate rejection rate than Box’s approximation when H_0 is true, but is slightly less accurate when H_0 is false.

IX. DATA ANALYSIS

Here we apply our results to ocean current speed and direction time series recorded at a mooring in the Labrador Sea [4], [17], [18]. We associate the eastward (zonal) measurement of current speed with $\{X_t\}$ and the northward (meridional) measurement with $\{Y_t\}$ and thus obtain the complex-valued series from (2). Series were recorded at six depths, (110, 760, 1260, 1760, 2510 and 3476m). The series are labelled 1 to 6 with increasing depth. We used $N = 1600$ observations for the 6-vector-valued complex time series, with a sampling interval of $\Delta_t = 1$ hr. In the spectral analysis $K = 12$ sine tapers were applied. Since W_N in (15) is 0.004c/hr, the validity range $W_N \leq |f| \leq f_N - W_N$ for our statistical results for a finite- N sample is given by $0.004 \leq |f| \leq 0.496$ c/hr. There was no evidence to reject the Gaussian assumption for this data set [5].

Of great interest to oceanographers are deep ocean motions well away from boundaries, especially in the internal wave frequency band. We pay special attention to low frequencies $f \in [0.02, 0.14]$, in the internal wave band and near to the semi-diurnal tidal frequency. The so-called ‘inertial frequency’ is approximately 0.07c/hr for this latitude and purely clockwise rotation occurs at the inertial frequency in the Northern hemisphere, making a band centred around the inertial frequency particularly interesting to study for such complex-valued processes. The dominant semi-diurnal tide at around $f = 0.08$ c/hr was estimated and removed to avoid spectral leakage affecting estimation near the inertial frequency.

For this data $\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{6,t}]^T$. In order to use different depth-contiguous sets of series we shall use the shorthand

$$\mathbf{Z}_{m:m'} \stackrel{\text{def}}{=} [Z_{m,t}, \dots, Z_{m',t}]^T \text{ with } 1 \leq m < m' \leq 6.$$

A. Concentration of Canonical Coherencies

The degree of polarization of a single random vector measures the spread amongst the eigenvalues of its covariance matrix. A random vector is completely polarized/unpolarized if all of its energy is concentrated in one direction/equally distributed amongst all dimensions. This idea can be extended to the correlation between two random vectors by defining the correlation spread [27] which provides a single, normalized measure of how much of the overall correlation is concentrated in a few coefficients, i.e., correlation is contained in a low dimensional subspace.

Using the analogous definition to [27] in our context we have *coherence spread* defined by

$$\sigma_p^2(f) \stackrel{\text{def}}{=} \frac{p}{p-1} \left(\frac{\sum_{i=1}^p l_j^4(f)}{(\sum_{i=1}^p l_j^2(f))^2} - \frac{1}{p} \right). \quad (58)$$

If only one canonical coherence is non-zero, then $\sigma_p^2(f) = 1$, whereas if all canonical coherencies are equal, $\sigma_p^2(f) = 0$. We note that if for a given f , $\sigma_p^2(f) = 1$, the likelihood ratio test statistic $T(f) = \prod_{j=1}^p (1 - l_j^2(f)) = 0$, i.e. achieves its minimum value, implying $\mathbf{R}_Z(f) \neq \mathbf{0}$. Of course, for $\sigma_p^2(f) < 1$ we are not able to conclude anything. In practice, we can only obtain an estimate $\hat{\sigma}_p^2(f)$ — where the $l_j^2(f)$ are replaced by the $\ell_j^2(f)$ — and therefore, the hypothesis test must be used to check for $\mathbf{R}_Z(f) = \mathbf{0}$.

Fig. 2 displays $\hat{\sigma}_p^2(f)$ for vector time series (a) $\mathbf{Z}_{1:2}$, (b) $\mathbf{Z}_{1:3}$, (c) $\mathbf{Z}_{1:4}$, (d) $\mathbf{Z}_{1:5}$ and (e) $\mathbf{Z}_{1:6}$. An immediate observation is that the coherence spread estimate for $\mathbf{Z}_{1:2}$ is highly erratic, with many values close to one. This is in contrast to all other plots where the spread ranges from 0.15 – 0.8 gradually decreasing in range as we consider time series at increasing depths. A notable feature of (b) $\mathbf{Z}_{1:3}$ is the broader peaks around 0.05, 0.065 and 0.11 and we see how the spread changes as we go from (b) $\mathbf{Z}_{1:3}$ to (c) $\mathbf{Z}_{1:4}$ with the broader peaks at 0.05 and 0.11 remaining intact whereas the one at 0.065 shrinks from its value of 0.7 to 0.4; the sharper peaks at 0.08, 0.09 and 0.138 disappear and a new peak appears at 0.044 which persists in both (d) $\mathbf{Z}_{1:5}$ and (e) $\mathbf{Z}_{1:6}$. We have thus seen how an additional series (depth) notably changes the concentration level of the overall coherence at some frequencies while disturbing it much less at others.

B. Test for Propriety

As defined in Section II-B the process $\{\mathbf{Z}_t\}$ is proper when $\mathbf{R}_Z(f) = \mathbf{0}$ for all $|f| \leq f_N$. Our test for $H_0 : \mathbf{R}_Z(f) = \mathbf{0}$ is valid, and may be carried out, for any $W_N < f < f_N - W_N$.

We test the same sets of time series for propriety and the results are displayed in Fig. 3. The solid line shows the test statistic $M(f)$ and the dotted line shows the critical value for each case. The test rejects H_0 at frequencies where $M(f)$ exceeds the critical value (thick line portions). The dashed line is the semi-diurnal tidal frequency. The coherence spread for $\mathbf{Z}_{1:2}$ (first subplot in Fig. 2) takes the maximum value of

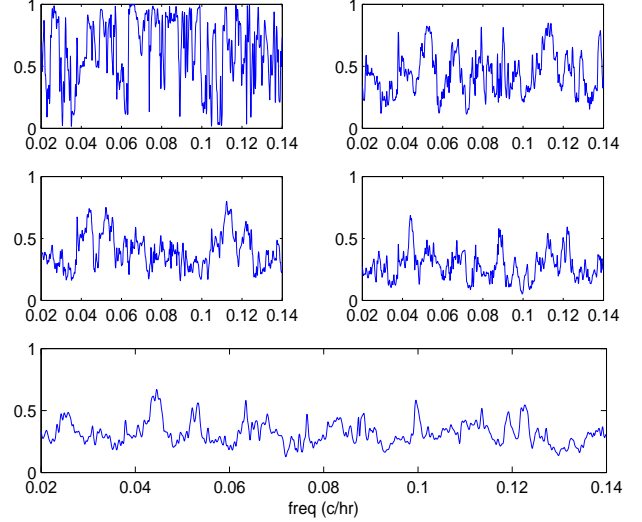


Fig. 2. Coherence spread estimate $\hat{\sigma}_p^2(f)$ for time series vectors (left to right, top to bottom): (a) $\mathbf{Z}_{1:2}$, (b) $\mathbf{Z}_{1:3}$, (c) $\mathbf{Z}_{1:4}$, (d) $\mathbf{Z}_{1:5}$ and (e) $\mathbf{Z}_{1:6}$.

0.9927 at $f = 0.065$, very close to the inertial frequency, and Fig. 3 (a) shows that our test rejects H_0 around this frequency very clearly. The band of frequencies around 0.04 is most prominent with rejection also clearly visible at frequencies 0.027, 0.075 and 0.087. For $\mathbf{Z}_{1:3}$, the test rejects H_0 for almost the same set of low frequencies with rejection also at a higher frequency around 0.12. Results for $\mathbf{Z}_{1:4}$ are very similar to that for $\mathbf{Z}_{1:3}$, the main difference being that a small frequency band near 0.1 also rejects H_0 . In general, we see that as other series (deeper in the ocean) are considered, H_0 is rejected, not only at low frequencies but also due to some additional higher frequencies, but less definitively so. Importantly then, Fig. 3 shows *which* frequency bands cause propriety to be rejected.

X. OTHER MEASURES OF VECTOR COHERENCE

From Lemma 1, one measure for vector coherence is the sum of all the canonical coherencies:

$$\text{tr}\{\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\} = \sum_{j=1}^p l_j^2(f). \quad (59)$$

Levikov and Sokolov [16] looked for a coefficient of coherence in the case of two real-valued vector random processes. In our paper, for the vector $\mathbf{V}_t = [\mathbf{X}_t^T, \mathbf{Y}_t^T]^T$, we consider $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ to be two geometrically related vector components and combine them to form a complex-valued vector time series. Levikov and Sokolov did not consider the two processes to be related in such a way and treated them simply as two vector process. They did, however, make use of the frequency domain and derived the quantity

$$\beta^2(f) \stackrel{\text{def}}{=} \frac{1}{2}[\mathbf{P}(f)\mathbf{S}_{\mathbf{Y}\mathbf{Y}}^{-1}(f) + \mathbf{S}_{\mathbf{Y}\mathbf{Y}}^{-1}(f)\mathbf{P}(f)],$$

where $\mathbf{P}(f) = \mathbf{S}_{\mathbf{Y}\mathbf{X}}(f)\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}(f)\mathbf{S}_{\mathbf{Y}\mathbf{X}}^H(f)$. Taking the trace of this quantity we get

$$\text{tr}\{\beta^2(f)\} = \frac{1}{2}\text{tr}\{\mathbf{P}(f)\mathbf{S}_{\mathbf{Y}\mathbf{Y}}^{-1}(f) + \mathbf{S}_{\mathbf{Y}\mathbf{Y}}^{-1}(f)\mathbf{P}(f)\}$$

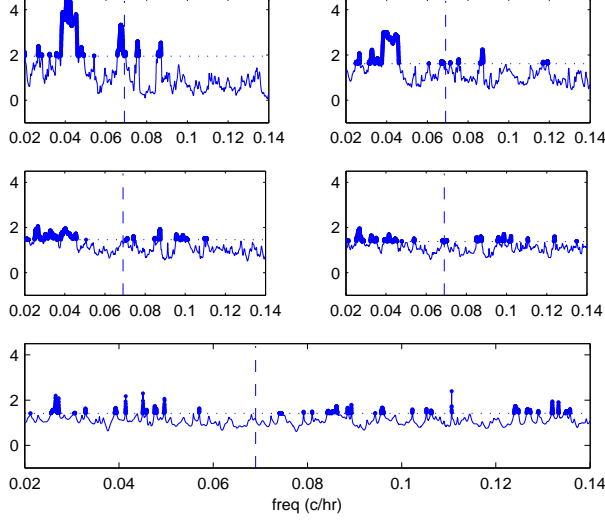


Fig. 3. The test statistic $M(f)$ (solid) and the critical value (dotted line) for (left to right, top to bottom): (a) $\mathbf{Z}_{1:2}$, (b) $\mathbf{Z}_{1:3}$, (c) $\mathbf{Z}_{1:4}$, (d) $\mathbf{Z}_{1:5}$ and (e) $\mathbf{Z}_{1:6}$. The test rejects the null hypothesis of propriety at frequencies where $M(f)$ exceeds the critical value (thick line portions). The dashed line is the semi-diurnal tidal frequency.

$$\begin{aligned}
 &= \text{tr}\{\mathbf{P}(f)\mathbf{S}_{\mathbf{Y}\mathbf{Y}}^{-1}(f)\} \\
 &= \text{tr}\{\mathbf{S}_{\mathbf{Y}\mathbf{X}}(f)\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}(f)\mathbf{S}_{\mathbf{Y}\mathbf{X}}^H(f)\mathbf{S}_{\mathbf{Y}\mathbf{Y}}^{-1}(f)\} \\
 &= \text{tr}\{\mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1}(f)\mathbf{S}_{\mathbf{X}\mathbf{Y}}(f)\mathbf{S}_{\mathbf{Y}\mathbf{Y}}^{-1}(f)\mathbf{S}_{\mathbf{Y}\mathbf{X}}(f)\}. \quad (60)
 \end{aligned}$$

This is of the same form as (59) only now using the components of the partition in (5); it will be the sum of all the canonical coherencies between $d\mathbf{Z}_X(f)$ and $d\mathbf{Z}_Y(f)$ where

$$\mathbf{X}_t = \int_{-f_N}^{f_N} e^{i2\pi ft} d\mathbf{Z}_X(f); \quad \mathbf{Y}_t = \int_{-f_N}^{f_N} e^{i2\pi ft} d\mathbf{Z}_Y(f).$$

Now $\mathbf{Z}_t = \mathbf{X}_t + i\mathbf{Y}_t$ and $\mathbf{Z}_t^* = \mathbf{X}_t - i\mathbf{Y}_t$. The spectral representation gives $d\mathbf{Z}(f) = d\mathbf{Z}_X(f) + id\mathbf{Z}_Y(f)$ and $d\mathbf{Z}^*(-f) = d\mathbf{Z}_X(f) - id\mathbf{Z}_Y(f)$. So we can write

$$\begin{bmatrix} d\mathbf{Z}(f) \\ d\mathbf{Z}^*(-f) \end{bmatrix} = \mathbf{T} \begin{bmatrix} d\mathbf{Z}_X(f) \\ d\mathbf{Z}_Y(f) \end{bmatrix}, \quad (61)$$

where \mathbf{T} is given in (3). We know that affine transformations of $d\mathbf{Z}_X(f)$ and of $d\mathbf{Z}_Y(f)$ will not change the canonical coherencies; however, (61) does not represent affine transforms of $d\mathbf{Z}_X(f)$ and of $d\mathbf{Z}_Y(f)$ since a mixing is involved.

Hence the quantities (59) and (60) will in general be different. Indeed for the example of Section VIII the value of (60) is zero over $|f| \leq 1/2$, because $\mathbf{S}_{\mathbf{X}\mathbf{Y}}(f) = \mathbf{S}_{\mathbf{Y}\mathbf{X}}(f) = \mathbf{0}$.

XI. SUMMARY AND CONCLUSION

We have developed a frequency domain approach to test for propriety of complex-valued vector time series. For propriety of $\{\mathbf{Z}_t\}$ we require $\mathbf{R}_Z(f) = \mathbf{0}$ for all $|f| \leq f_N$. We can carry out the test $H_0 : \mathbf{R}_Z(f) = \mathbf{0}$ for any $W_N < f < f_N - W_N$. Most importantly for the vector case ($p \geq 2$) we have justified use of the rule that H_0 is rejected if $M(f) = -2K \log T(f) > bF_{\nu_1, \nu_2}(1 - \alpha)$. There is no assumption that K is large, and indeed this would rarely be

expected in practice. We have shown in detail how the statistic $T(f)$ arises by consideration of canonical coherencies for complex-valued vector time series. When propriety is invalid, the frequency domain approach has the scientific advantage of showing which frequency bands are causing rejection, likely allowing linkage to known or hypothesized properties of the physical processes involved.

APPENDIX

A. Proof of Lemma 1

Given (21), since $\mathbf{S}_Z(f)$ and $\mathbf{S}_Z^T(-f)$ are positive-definite (Hermitian) covariance matrices, we have solutions

$$\mathbf{A}^H(f) = \mathbf{F}^H(f)\mathbf{S}_Z^{-1/2}(f) \quad (62)$$

$$\mathbf{B}^H(f) = \mathbf{G}^H(f)\mathbf{S}_Z^{-T/2}(-f) \quad (63)$$

where $\mathbf{F}(f), \mathbf{G}(f) \in \mathbb{C}^{p \times p}$ are unitary. Then

$$\begin{aligned}
 \mathbf{K}(f) &= \mathbf{A}^H(f)\mathbf{R}_Z(f)\mathbf{B}(f) \\
 &= \mathbf{F}^H(f)\mathbf{S}_Z^{-1/2}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T/2}(-f)\mathbf{G}(f). \quad (64)
 \end{aligned}$$

We now make use of the weak majorization result [28, p. 294]. Let $|\text{diag}(\mathbf{K}(f))| \stackrel{\text{def}}{=} [|K_{11}(f)|, \dots, |K_{pp}(f)|]^T$. Then

$$\sum_{j=1}^r |K_{jj}(f)| \leq \sum_{j=1}^r |\sigma_j(f)|, \quad r = 1, \dots, p.$$

where the $\sigma_j(f)$ are the singular values of $\mathbf{K}(f)$ and $\sigma_{[1]}(f) \geq \sigma_{[2]}(f) \geq \dots \geq \sigma_{[p]}(f)$, (a descending size order). Hence the solution to (23) is found by making $\mathbf{K}(f)$ diagonal. From (64), we thus choose $\mathbf{F}(f)$ and $\mathbf{G}(f)$ to diagonalize $\mathbf{K}(f)$, i.e., $\mathbf{F}(f)$ and $\mathbf{G}(f)$ are determined by singular value decomposition of

$$\mathbf{C}(f) \stackrel{\text{def}}{=} \mathbf{S}_Z^{-1/2}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T/2}(-f) = \mathbf{F}(f)\mathbf{K}(f)\mathbf{G}^H(f),$$

giving

$$\begin{aligned}
 \mathbf{C}(f)\mathbf{C}^H(f) &= \mathbf{S}_Z^{-1/2}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{S}_Z^{-1/2}(f) \\
 &= \mathbf{F}(f)\mathbf{L}(f)\mathbf{F}^H(f),
 \end{aligned}$$

where $\mathbf{L}(f)$ denotes a diagonal matrix with j th element $l_j^2(f) = |K_{jj}(f)|^2$, in descending size order. Now multiply through on the left by $\mathbf{S}_Z^{-1/2}(f)$ and on the right by $\mathbf{F}(f)$ to obtain

$$\begin{aligned}
 \mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{S}_Z^{-1/2}(f)\mathbf{F}(f) \\
 = \mathbf{S}_Z^{-1/2}(f)\mathbf{F}(f)\mathbf{L}(f),
 \end{aligned}$$

which, using (62), can be written

$$\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{A}(f) = \mathbf{A}(f)\mathbf{L}(f),$$

so that

$$\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{A}_j(f) = l_j^2(f)\mathbf{A}_j(f),$$

and $l_j^2(f)$ are the eigenvalues, and $\mathbf{A}_j(f)$ are the eigenvectors of the $p \times p$ matrix $\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)$, as required. Note that this matrix is the product of the two Hermitian matrices $\mathbf{S}_Z^{-1}(f)$ and $\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)$.

Similarly,

$$\begin{aligned} \mathbf{C}^H(f)\mathbf{C}(f) &= \mathbf{S}_Z^{-T/2}(-f)\mathbf{R}_Z^H(f)\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T/2}(-f) \\ &= \mathbf{G}(f)\mathbf{L}(f)\mathbf{G}^H(f). \end{aligned}$$

Multiply through on the left by $\mathbf{S}_Z^{-T/2}(f)$ and on the right by $\mathbf{G}(f)$, and use (63) to obtain

$$\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{B}(f) = \mathbf{B}(f)\mathbf{L}(f),$$

so that, as required

$$\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{B}_j(f) = l_j^2(f)\mathbf{B}_j(f).$$

$l_j^2(f)$ are the eigenvalues, and $\mathbf{B}_j(f)$ are the eigenvectors of the $p \times p$ matrix $\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)$.

Notice that the matrices $\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)$ and $\mathbf{S}_Z^{-T}(-f)\mathbf{R}_Z^H(f)\mathbf{S}_Z^{-1}(f)\mathbf{R}_Z(f)$ are just cyclic permutations of each other.

Finally,

$$\text{cov}\{\mathbf{dZ}_\xi(f), \mathbf{dZ}_\eta(f)\} = E\{\mathbf{dZ}_\xi(f)\mathbf{dZ}_\eta^H(f)\} = \mathbf{K}(f),$$

and the solution of the optimization problem makes $\mathbf{K}(f)$ diagonal. Hence, $\text{corr}\{\mathbf{dZ}_{\xi_j}(f), \mathbf{dZ}_{\eta_k}(f)\} = 0$, for $j, k = 1, \dots, p; j \neq k$.

B. Proof of Lemma 2

To simplify notation we drop explicit frequency dependence. Consider the distribution of $T = l_G^{1/K}$, given by (38), under the null hypothesis. We have

$$E\{T^r\} = \int \cdots \int \frac{[\det\{\mathbf{A}\}]^r}{[\det\{\mathbf{A}_{11}\}\det\{\mathbf{A}_{22}\}]^r} g(\mathbf{A}) d\mathbf{A},$$

where $g(\mathbf{A})$ is the PDF for the complex Wishart distribution. As explained in the text under the null hypothesis we can take \mathbf{S}_U to be \mathbf{I}_{2p} because of invariance under the group action. We can thus replace (36) by

$$\mathbf{A} \triangleq \mathcal{W}_{2p}^C(K, \mathbf{I}_{2p}),$$

and using [11] we know that for $\mathbf{A} > 0, K \geq 2p$,

$$g(\mathbf{A}; K, 2p, \mathbf{I}_{2p}) = c(K, 2p)[\det\{\mathbf{A}\}]^{K-2p} e^{-\text{tr}\{\mathbf{A}\}}, \quad (65)$$

where $c(K, 2p)$ is a constant defined by

$$c^{-1}(K, 2p) = \pi^{p(2p-1)} \prod_{i=1}^{2p} \Gamma(K+1-i). \quad (66)$$

So $E\{T^r\}$ takes the form

$$\begin{aligned} &c(K, 2p) \int \cdots \int \frac{[\det\{\mathbf{A}\}]^{K-2p+r}}{[\det\{\mathbf{A}_{11}\}\det\{\mathbf{A}_{22}\}]^r} e^{-\text{tr}\{\mathbf{A}\}} d\mathbf{A} \\ &= \frac{c(K, 2p)}{c(K+r, 2p)} \int \cdots \int \frac{1}{[\det\{\mathbf{A}_{11}\}\det\{\mathbf{A}_{22}\}]^r} \\ &\quad \times [c(K+r, 2p)[\det\{\mathbf{A}\}]^{K-2p+r} e^{-\text{tr}\{\mathbf{A}\}}] d\mathbf{A}. \end{aligned}$$

The integration is w.r.t. $d\mathbf{A} = dA_{11}, \dots, dA_{2p2p}$. The term in the square brackets above is the PDF for the $\mathcal{W}_{2p}^C(K+r, \mathbf{I}_{2p})$ distribution. The integral of this density with respect to the

elements in \mathbf{A}_{12} and \mathbf{A}_{21} must give the marginal density of $\mathbf{A}_{11}, \mathbf{A}_{22}$, which is the product

$$g(\mathbf{A}_{11}; K+r, p, \mathbf{I}_p) \cdot g(\mathbf{A}_{22}; K+r, p, \mathbf{I}_p), \quad (67)$$

since \mathbf{A}_{11} and \mathbf{A}_{22} are independent under the null hypothesis. Carrying out the integration and using (65) and (67) we obtain

$$\begin{aligned} &\frac{c(K, 2p)}{c(K+r, 2p)} \int \cdots \int \frac{1}{[\det\{\mathbf{A}_{11}\}\det\{\mathbf{A}_{22}\}]^r} \\ &\quad \times \prod_{j=1}^2 c(K+r, p) [\det\{\mathbf{A}_{jj}\}]^{K+r-p} e^{-\text{tr}\{\mathbf{A}_{jj}\}} d\mathbf{A}_{jj} \\ &= \frac{c(K, 2p)}{c(K+r, 2p)} \prod_{j=1}^2 \int \cdots \int c(K+r, p) [\det\{\mathbf{A}_{jj}\}]^{K-p} \\ &\quad \times e^{-\text{tr}\{\mathbf{A}_{jj}\}} d\mathbf{A}_{jj} \\ &= \frac{c(K, 2p)}{c(K+r, 2p)} \prod_{j=1}^2 \frac{c(K+r, p)}{c(K, p)}. \end{aligned} \quad (68)$$

Using (66) and (68) we get

$$E\{T^r\} = \frac{\prod_{j=1}^{2p} \Gamma(K+r+1-j) [\prod_{j=1}^p \Gamma(K+1-j)]^2}{\prod_{j=1}^{2p} \Gamma(K+1-j) [\prod_{j=1}^p \Gamma(K+r+1-j)]^2}. \quad (69)$$

This agrees with [14, eqn. (2.6)] which appears without reference or proof. Now $T^r = l_G^{r/K}$ so if we let $r \rightarrow rK$, then $T^{rK} = l_G^r$. So

$$\begin{aligned} E\{l_G^r\} &= \frac{\prod_{j=1}^{2p} \Gamma(K[1+r]+1-j) \prod_{j=1}^p \Gamma(K+1-j)}{\prod_{j=1}^{2p} \Gamma(K+1-j) \prod_{j=1}^p \Gamma(K[1+r]+1-j)} \\ &\quad \times \left[\frac{\prod_{j=1}^p \Gamma(K+1-j)}{\prod_{j=1}^p \Gamma(K[1+r]+1-j)} \right] \\ &= \frac{\prod_{j=1}^p \Gamma(K[1+r]+1-j-p)}{\prod_{j=1}^p \Gamma(K+1-j-p)} \\ &\quad \times \left[\frac{\prod_{j=1}^p \Gamma(K+1-j)}{\prod_{j=1}^p \Gamma(K[1+r]+1-j)} \right] \\ &= \left[\frac{\prod_{j=1}^p \Gamma(K+1-j)}{\prod_{j=1}^p \Gamma(K+1-j-p)} \right] \\ &\quad \times \frac{\prod_{j=1}^p \Gamma(K[1+r]+1-j-p)}{\prod_{j=1}^p \Gamma(K[1+r]+1-j)}, \end{aligned}$$

which is (39).

C. Proof of Lemma 3

Under the null hypothesis the r th moment of $T(f)$ is

$$E\{T^r(f)\} = \prod_{j=1}^p \frac{\Gamma(K+r+1-j-p)\Gamma(K+1-j)}{\Gamma(K+r+1-j)\Gamma(K+1-j-p)}. \quad (70)$$

To see this start with (69) and proceed in analogous vein to the last part of the proof of Lemma 2; since we are continuing to look at $E\{T^r\}$ the step $r \rightarrow rK$ is not made. Note that when $j = p$ the critical gamma function argument is still positive: $K+r+1-j-p = K+r+1-2p > 0$ since $K \geq 2p$ with $r \geq 0$.

A real scalar random variable X is said to have a (type-1) beta distribution, $X \stackrel{d}{=} \text{beta}(\alpha, \beta)$, if the PDF is

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1, \alpha > 0, \beta > 0.$$

The r th moment for this distribution is

$$E\{X^r\} = \frac{\Gamma(\alpha + r)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + r)\Gamma(\alpha)}, \quad \alpha + r > 0. \quad (71)$$

Comparing (70) and (71) we see for a fixed j that $\alpha = K + 1 - j - p$ and $\beta = p$ which gives the required result.

ACKNOWLEDGEMENT

The authors are grateful to Jon Lilly for the Labrador Sea data.

REFERENCES

- [1] T. Adali, P. J. Schreier & L. L. Scharf, "Complex-valued signal processing: the proper way to deal with impropriety," *IEEE Transactions on Signal Processing*, vol. 59, pp. 5101–5125, 2011.
- [2] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, 317–46, 1949.
- [3] E. M. Carter, C. G. Khatri & M. S. Srivastava, "Nonnull distribution of likelihood ratio criterion for reality of covariance matrix," *J. Multivariate Analysis*, vol. 6, pp. 176–184, 1976.
- [4] S. Chandna and A. T. Walden, "Statistical properties of the estimator of the rotary coefficient," *IEEE Transactions on Signal Processing*, vol. 59, pp. 1298–1303, 2011.
- [5] S. Chandna and A. T. Walden, "Simulation methodology for inference on physical parameters of complex vector-valued signals," *IEEE Transactions on Signal Processing*, vol. 61, pp. 5260–5269, 2013.
- [6] T. Chonavel, *Statistical Signal Processing*. London UK: Springer-Verlag, 2002.
- [7] S. Elipot and R. Lumpkin, "Spectral description of oceanic near-surface variability," *Geophys. Res. Lett.* **35**, L05606, 2008.
- [8] W. J. Emery and R. E. Thomson, *Data Analysis Methods in Physical Oceanography*. New York: Pergamon, 1998.
- [9] C. Fang, P. R. Krishnaiah and B. N. Nagarsenker, "Asymptotic distributions of the likelihood ratio test statistics for covariance structures of the complex multivariate normal distributions," *J. Multivariate Analysis*, vol. 12, pp. 597–611, 1982.
- [10] P. Ginzberg, "Quaternion matrices: statistical properties and applications to signal processing and wavelets," Ph. D. dissertation, Dept. Mathematics, Imperial College London, 2013.
- [11] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152–77, 1963.
- [12] A. K. Gupta, "Distribution of Wilks' likelihood-ratio criterion in the complex case," *Ann. Instit. Statist. Math.*, vol. 23, pp. 77–87, 1971.
- [13] A. K. Gupta, "On a test for reality of the covariance matrix in a complex Gaussian distribution," *J. Statist. Computation and Simulation*, vol. 2, pp. 333–342, 1973.
- [14] P. R. Krishnaiah, J. C. Lee and T. C. Chang, "The distributions of the likelihood ratio statistics for tests of certain covariance structures of complex multivariate normal populations," *Biometrika*, vol. 63, pp. 543–549, 1976.
- [15] P. R. Krishnaiah, J. C. Lee and T. C. Chang, "Likelihood ratio tests on covariance matrices and mean vectors of complex multivariate normal populations and their applications in time series," Technical report TR-83-03, Center for Multivariate Analysis, Univ. Pittsburgh, 1983.
- [16] S. Levikov and S. Sokolov, "Coefficient of coherence in the case of two vector random processes," *Deep-Sea Research Part I: Oceanographic Research Papers*, vol. 44, pp. 1329–1338, 1997.
- [17] J. M. Lilly, P. B. Rhines, M. Visbeck, R. Davis, J. R. Lazier, F. Schott and D. Farmer, "Observing deep convection in the Labrador Sea during winter 1994/95," *J. Phys. Oceanogr.*, vol. 29, 2065–98, 1999.
- [18] J. M. Lilly and P. B. Rhines, "Coherent eddies in the Labrador Sea observed from a mooring," *J. Phys. Oceanogr.*, vol. 32, 585–98, 2002.
- [19] C. N. K. Mooers, "A technique for the cross spectrum analysis of pairs of complex-valued time series, with emphasis on properties of polarized components and rotational invariants," *Deep-Sea Research*, vol. 20, pp. 1129–1141, 1973.
- [20] W. Min and R. S. Tsay, "On canonical analysis of multivariate time series," *Statistica Sinica*, vol. 15, pp. 303–323, 2005.
- [21] M. Miyata, "Complex generalization of canonical correlation and its application to a sea-level study," *J. Marine Research*, vol. 28, pp. 202–214, 1970.
- [22] J. Navarro-Moreno, M. D. Estudillo-Martínez, R. M. Fernández-Alcalá & J. C. Ruiz-Molina, "Estimation of improper complex-valued random signals in colored noise by using the Hilbert space theory," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2859–2867, 2009.
- [23] E. Ollila and V. Koivunen, "Generalized complex elliptical distributions," in *Proc. Third Sensor Array and Multichannel Signal Processing Workshop*, Sitges, Spain, July, 2004, pp. 460–4.
- [24] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing* vol. 44, pp. 2637–2640, 1996.
- [25] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 411–420, 1997.
- [26] G. Reinsel, *Elements of Multivariate Time Series Analysis (2nd Ed)*. New York: Springer, 1997.
- [27] P. J. Schreier, "A unifying discussion of correlation analysis for complex random vectors," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1327–1336, 2008.
- [28] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge UK: Cambridge University Press, 2010.
- [29] P. J. Schreier, L. L. Scharf and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Process. Lett.*, vol. 13, pp. 433–6, 2006.
- [30] A. M. Sykulski, S. C. Olhede, J. M. Lilly & J. J. Early, "On parametric modelling and inference for complex-valued time series," ArXiv preprint arXiv:1306.5993v2, 2015. Available: <http://arxiv.org/abs/1306.5993>.
- [31] A. M. Sykulski, S. C. Olhede & J. M. Lilly, "An improper complex autoregressive process of order one," ArXiv preprint arXiv:1511.04128v1, 2015. Available: <http://arxiv.org/abs/1511.04128>.
- [32] H. van Haren and C. Millot, "Rectilinear and circular inertial motions in the Western Mediterranean Sea," *Deep-Sea Research Part I*, vol. 51, pp. 1441–55, 2004.
- [33] A. T. Walden, "Rotary components, random ellipses and polarization: a statistical perspective," *Phil. Trans. R. Soc. A*, vol. 371, doi:10.1098/rsta.2011.0554, 2013.
- [34] A. T. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Transactions on Signal Processing* vol. 57, pp. 825–834, 2009.
- [35] A. T. Walden, E. J. McCoy and D. B. Percival, "The effective bandwidth of a multitaper spectral estimator," *Biometrika*, vol. 82, 201–214, 1995.
- [36] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions, Volume I: Basic Results*. New York: Springer, 1987.
- [37] G. A. Young and R. L. Smith, *Essentials of Statistical Inference*. Cambridge UK: Cambridge University Press, 2005.